Project Description

Most data in cold-atom experiments comes from images, the analysis of which is limited by our preconceptions of the patterns that could be present in the data. In this project, we focus on the well-defined case of detecting dark solitons—appearing as local density depletions in a Bose-Einstein condensates (BECs)—using a methodology that is extensible to the general task of pattern recognition in images of cold atoms. Studying soliton dynamics over a wide range of parameters requires the analysis of large datasets, making the existing human-inspection-based methodology a significant bottleneck. To enable machine learning analysis, we established a dataset of over 6 000 labeled experimental images of BECs with and without dark solitonic excitations.

If you use this dataset, please cite our paper:
Shangjie Guo, Amilson R. Fritsch, Craig Greenberg, I. B. Spielman, and Justyna P. Zwolak, "Machine-learning enhanced dark soliton detection in Bose-Einstein condensates", arXiv:2101.05404.

Dark solitons in BECs dataset

This dataset consists 6 257 absorption images with human assigned labels. These images were taken from multiple experiments performed in a single lab over a span of two months. Each record of atomic density in the dataset was obtained by combining three raw images:

- the probe with the BEC's shadow $I^{\mathrm{A}}_{i,j}$ (see Fig. 1(a)),
- the probe intensity $I^{\mathrm{P}}_{i,j}$ (see Fig. 1(b)),
- a dark frame containing any ambient background signal $I^{\mathrm{BG}}_{i,j}$ (see Fig. 1(c)).

The raw images (Fig. 1(a-c)) were obtained with a 648×488 pixel camera (Point Grey FL3) with 5.6 µm square pixels, labeled by i and j. Including the $\approx 6\times$ magnification, each pixel has effective size of 0.93 µm. The diffraction limit of the imaging system gives an optical resolution of ≈ 2.8 µm (roughly three pixels).

The three raw images are combined to produce the 2D density using to the following relation:

$$\sigma_0 n_{i,j} \approx -\ln\left[\frac{I^{\mathrm{A}}_{i,j} - I^{\mathrm{BG}}_{i,j}}{I^{\mathrm{P}}_{i,j} - I^{\mathrm{BG}}_{i,j}}\right], \tag{1}$$

where the resonant cross-section $\sigma_0 = 3\lambda^2/(2\pi)$ is derived from the wavelength λ of the probe laser. The dimensionless product $\sigma_0 n_{i,j}$ is of order 1 in our data, so we express density in terms of this product.

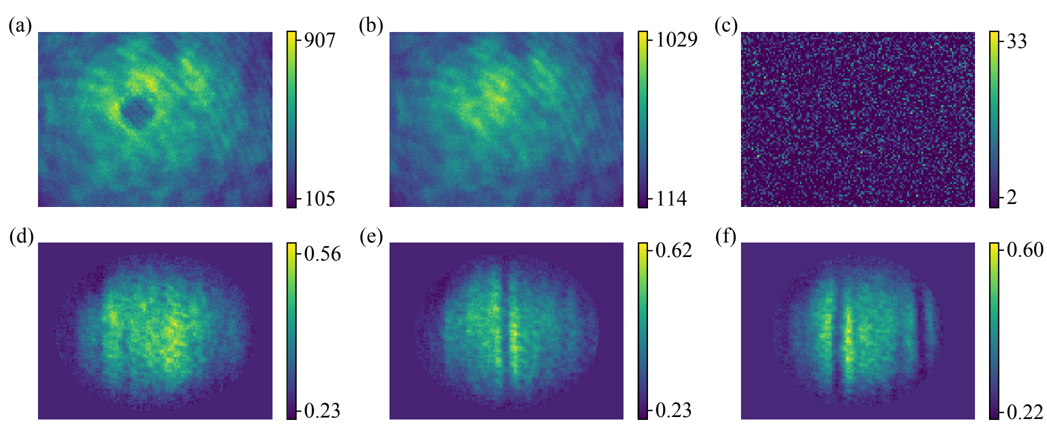

Figure 1: Sample data. (a-c) Raw absorption images: (a) atom I^{A}, (b) probe I^{P}, and (c) background I^{BG}. (d-f) Pre-processed images that labeled as (d) no soliton, (e) single soliton, and (f) other excitations.

As can be seen in (Fig. 1(a), the BEC occupies only a small region of the image, and the long axis of the BEC is rotated with respect to the camera. Therefore, the 2D absorption images images are rotated to align the BEC with the image frame and cropped to discard the large fraction of the image that does not contain information about the BEC. Since the BEC's position and shape can vary for different realizations of the same experiment, we implement a fitting approach to determine the position and size of the BEC.

We fit every image to a column-integrated 3D Thomas-Fermi distribution describing the density distribution of 3D BECs integrated along the imaging axis:

$$n_{i,j}^{\mathrm{TF}} = n_0 \max\left\{\left[1-\left(\frac{i-i_0}{R_i}\right)^2-\left(\frac{j-j_0}{R_j}\right)^2\right],0\right\}^{3/2}+\delta n. \tag{2}$$

We use six parameters to fit: the BEC center coordinates $[i_0, j_0]$; the peak 2D density n_0; the Thomas-Fermi radii $[R_i,\ R_j]$; and a offset δn from small changes in probe intensity between images.

We determined the 164×132 pixel extent of the cropping region by examining the radii $[R_i, R_j] = [66(5), 58(3)]$ obtained from fits to images included in this dataset. We then centered the cropping region at $[i_0, j_0]$ as determined from fits of each image separately. The process was validated on an additional 10^4 images not included in the dataset. Finally, an elliptical mask with radii [Ri,Rj] was applied to each image, eliminating all technical noise outside the BEC. In the resulting images dark solitons appear as vertically aligned density depletions and are easily visually identified (Fig. 1(d)).

Three human labelers labeled the preprocessed data, categorizing the images into three classes as: 0 (no soliton), 1 (single soliton), and 2 (other excitations). The "no soliton" class contains images that unambiguously contains no solitons; the "single soliton" class describes images with one and only one soliton; and "other excitations" class covers any image that can neither be interpreted as "no soliton" nor "single soliton." Three human labelers first label the images individually, then label the disagreed images by discussing with each other.

Data structure

Each absorption image in the dataset is stored as a dictionary in a separate NumPy file. The dictionary contains two elements (keys):

- `'data'` : the preprocessed absorption image [(132, 164) numpy.array],
- `'label'`: human assigned labels: 0, 1, or 2; [int],

where the type of each element in the dictionary is given in the brackets.